Ecology, 69(4), 1988, pp. 1266–1274

THE USE OF REGRESSION FOR DETECTING COMPETITION WITH MULTICOLLINEAR DATA[1]

BRUCE A. CARNES
Division of Biological and Medical Research, Argonne National Laboratory
Argonne, Illinois 60439-4833 USA

AND

NORMAN A. SLADE
Museum of Natural History and Department of Systematics and Ecology, University of Kansas,
Lawrence, Kansas 66045 USA

Abstract. Monte Carlo simulations were used to demonstrate that regression methods could be successfully used to estimate competition coefficients with collinear data, but only under conditions that may be difficult to meet with ecological data. Ordinary least squares performs well when estimating coefficients associated with noncollinear predictor variables. Stepwise regression and maximum eigenvalue least squares reduce collinearity by deleting information that, even though not statistically significant, may be important in accurately estimating interaction. We propose that apparent competition (Holt 1977) and the failure of regression techniques to detect competition when it is known to exist experimentally may be due to the omission or lack of measurement of critical elements of the community matrix.

Key words: collinearity; competition; regression; resource use; simulation.

INTRODUCTION

Interspecific competition plays a central role in ecological theory, and numerous approaches to quantifying competition from observational data have been proposed. Traditional approaches, based on resource overlap, use the methods of MacArthur and Levins (1967) and Levins (1968). An alternative approach that uses linear regression was proposed by Schoener (1974) to analyze fluctuations around equilibrium in homogeneous environments. Seifert and Seifert (1976) used the partial regression coefficient from a series of population estimates for two species over time as an interaction coefficient to be interpreted as indicating competition or some other symbiotic relationship. Eventually, this approach was modified (Crowell and Pimm 1976) to account for effects due to habitat heterogeneity and applied to field data. The regression approach to estimating competition coefficients was demonstrated (Hallett and Pimm 1979) by Monte Carlo methods to be appropriate even in the presence of random deviations in population density away from equilibrium. It appears that the regression approach has been generally accepted and frequently applied (Dueser and Hallett 1980, Hallett 1982, Porter and Dueser 1982, Crowell 1983, Hallett et al. 1983). However, on the basis of their inability to detect competition by regression methodology when competition was known to exist experimentally, Abramsky, Rosenzweig, and associates (Rosenzweig et al. 1984, 1985, Abramsky et al. 1985, 1986) have questioned the efficacy of regression for estimating interaction coefficients, and Pimm (1985) and Schoener (1985) have published replies.

Because ecology is the study of interrelationships, the predictor variables used in a regression model applied to ecological data are rarely, if ever, truly independent. In the statistical literature, this problem is known as multicollinearity (Farrar and Glauber 1967, Mason et al. 1975). Multicollinearity results in regression coefficients that are potentially unstable (Silvey 1969, McDonald and Schwing 1973), have inflated variances (Marquardt 1970, Greenberg 1975), and may deviate significantly from the true parameter values (Mason et al. 1975). Because multicollinearity is inherent in ecological data, we were highly skeptical of the reliability of the competition coefficients derived from the regression approach. Therefore, we decided to test the reliability of regression analysis to correctly estimate interaction coefficients by using Monte Carlo simulation.

METHODS

Monte Carlo simulation was used to generate data sets with carefully defined characteristics for analysis. For simplicity, only two species of consumers were considered; however, our approach can just as easily be applied to more complex systems. Initially, we restricted our analysis to data representing a single time period and four resource dimensions.

For each set of simulations, a 6 × 6 parametric cor-

[1] Manuscript received 26 June 1987; revised 27 November 1987; accepted 16 December 1987.

TABLE 1. Parametric correlation structure among consumers (C1, C2) and resource dimensions (R1, R2, R3, R4) used to generate the competition (C), neutralism (N), and mutualism (M) simulations.

						C2	
	C1	R1	R2	R3	R4	C	N	M
C1	1.00	0.48	0.38	−0.58	0.37	0.50	0.00	0.50
R1		1.00	0.55	0.09	0.79	0.45	0.00	0.00
R2			1.00	0.39	0.93	0.36	0.00	0.00
R3				1.00	0.43	−0.52	0.00	−0.08
R4					1.00	0.34	0.00	0.00
						1.00	1.00	1.00

relation matrix (R) was chosen to express the interrelationships between the two consumers and among the consumers and the resources (Table 1). A 4 × 4 submatrix involving the correlations among resources was chosen to represent correlations that could occur in an actual field study (Carnes 1980). This submatrix of the parametric correlation matrix was invariant for all simulations. In addition, the correlations between the first consumer and the resource dimensions were invariant for all simulations. Only the correlations between the two consumers and between the second consumer and the resource dimensions were allowed to vary between simulations (Table 1). Many correlation structures were analyzed but only three will be presented. These three correlation structures were chosen to reflect varying degrees of correlation between the second consumer and the resources and a range of possible relationships between the two consumers (competition, mutualism, neutralism: Odum 1971).

After the 6 × 6 parametric correlation matrix was determined, the eigenvalues and eigenvectors were extracted (Green and Carroll 1976). The eigenvector matrix (E) and the diagonal matrix of eigenvalues [D(λ)] can be used to reconstitute the parametric correlation matrix:

$$R = ED(\lambda)E'. \quad (1)$$

Using this relationship, we calculated a 6 × 6 matrix $F = ED(\sqrt{\lambda})$ such that $FF' = R$. Next, we generated a 50 × 6 matrix (Z) of independent random standard normal deviates. Finally, Z was postmultiplied by F′ yielding a 50 × 6 data matrix ($X = ZF'$) that, when premultiplied by its transpose, gave an approximation to the parametric correlation matrix (R). The matrix (X) simulated the results of 50 independent censuses, drawn either from 50 sites or from widely spaced time periods so autocorrelations between censuses were negligible, and the complications discussed by Wilson (1985) were avoided.

Each sample correlation matrix, $FZ'ZF' = X'X$, deviated from R because sample correlations among the six sets of 50 independent random normal deviates were not actually zero. One "consumer" (first or last column of the data matrix, X) was selected as the response vector (Y) and three different regression techniques were used to estimate the interaction coefficients. The first approach was ordinary least squares (OLS) using the remaining five variables from the X matrix as predictors. Next, maximum eigenvalue least squares (MELS) (Massy 1965, Hawkins 1973, Gunst and Mason 1977, Ginevan and Carnes 1981) was used because it is specifically designed for situations involving multicollinearity. Finally, we used the modified stepwise (SW) regression approach described by Crowell and Pimm (1976). The entire process was then repeated with the second consumer species used as the response variable and the first species used as one of the five predictors. We ran 100 replications for each consumer selected as the response variable and for each of the three R matrices (i.e., competition, neutralism, and mutualism).

Means, variances, and standard errors of the estimated regression coefficient for the consumer predictor variable (i.e., the coefficient representing consumer relationship) were calculated for each regression technique. The vector of parametric correlations for the consumer response variable with the predictor variables (R_{XY}) and the matrix of parametric correlations among the predictor variables (R_{XX}) extracted from R were used to generate the parametric regression coefficients ($B = R_{XX}^{-1}R_{XY}$) for comparison with the mean values of the simulation consumer coefficients. Percentiles of the distribution of estimated regression coefficients were scanned for indications of the frequency of type I or type II errors in detecting interspecific interactions. The sensitivity or stability of the three regression approaches to collinearity were evaluated by comparing the condition indices (Belsley et al. 1980) of predictor variable correlation matrices. Finally, we reran all simulations, including 15 additional independent nuisance variables to simulate the inclusion of extraneous habitat variables. All analyses were conducted with the PROC MATRIX procedure (SAS 1985) on the IBM 3033 at Argonne National Laboratory.

Three approaches were used to check the validity of the simulation technique. Common factor analysis, PROC FACTOR (SAS 1985), was applied to each of the three 6 × 6 parametric correlation matrices (R) to estimate communalities (Afifi and Clark 1984). The communalities for the two consumer variables represented the parametric coefficient of determination (R^2) to be expected from the OLS solutions. Next, 95%

TABLE 2. Simulation summary without nuisance variables. OLS = ordinary least squares; MELS = maximum eigenvalue least squares; SW = stepwise regression.

	Response variable					
	Consumer 1			Consumer 2		
Statistics*	OLS	MELS	SW	OLS	MELS	SW
			Competition			
Parametric	−0.758	−0.758	−0.758	−1.259	−1.259	−1.259
Estimate	−0.765	−0.744	−0.670	−1.270	−1.171	−1.148
SE	0.010	0.010	0.010	0.017	0.040	0.024
C.Index	761.2	19.6	30.1	740.0	32.3	123.6
$\|Z\|$	0.66	1.45	8.5	0.63	2.20	4.65
P	.51	.15	<.001	.53	.03	<.001
			Neutralism			
Parametric	0.000	0.000	0.000	0.000	0.000	0.000
Estimate	0.006	0.006	0.006	0.035	0.025	−0.001
SE	0.006	0.006	0.006	0.032	0.030	0.012
C.Index	716.3	8.2	29.7	229.3	32.2	2.7
$\|Z\|$	0.94	1.05	1.00	1.08	0.85	0.04
P	.35	.30	.32	.28	.40	.96
			Mutualism			
Parametric	0.438	0.438	0.438	2.276	2.276	2.276
Estimate	0.436	0.433	0.423	2.313	2.174	0.560
SE	0.005	0.005	0.005	0.025	0.052	0.027
C.Index	705.9	8.4	20.8	735.4	34.2	2.4
$\|Z\|$	0.40	1.11	3.08	1.46	1.97	64.30
P	.69	.27	.002	.14	.05	<.001

* Expected value from parameteric correlation matrix; SE = standard error of the estimate; C.Index = condition index of $(x'x)^{-1}$; $|Z|$ = absolute value of standard normal statistic; P = significance level of Z.

confidence intervals were calculated for the nondiagonal elements of the simulated correlation matrices to determine whether they included the parametric correlations. Finally, Fisher's Z transformation was applied to the nondiagonal elements of the simulated correlation matrices ($x'x$) and their variance was compared with an expectation of $1/(N-3)$ (Sokal and Rohlf 1969).

RESULTS

None of the three approaches that were used to check the validity of the simulation technique uncovered anything unusual. The communalities generated for the two consumer variables from the common factor analysis were in close agreement with the observed coefficients of multiple determination for the ordinary least squares (OLS) solutions. Confidence intervals for the elements of the sample correlation matrices did include the parametric values, and their variance was in agreement with expectation.

The analyses of three sets of Monte-Carlo–simulated data illustrate the general patterns found in the entire set of results (Table 2). A condition index <10 represents a stable matrix, and condition indices from 30 to 100 suggest moderate to strong dependencies (Belsley et al. 1980). Condition indices >700 for the OLS solutions suggest that extreme collinearity exists in the data. Maximum eigenvalue least squares (MELS), which deletes eigenvectors (not variables) associated with the smallest eigenvalues, and stepwise regression (SW), which prevents correlated variables from entering the model, both greatly improve (i.e., decrease) the condition index (Table 2).

Despite collinearity problems, the mean OLS estimates do not differ significantly from the known parametric values (listed in Table 2). MELS estimates, though in good agreement with expectation, differ significantly from the parametric values in two cases (consumer 2, competition and mutualism). The modified stepwise approach of Crowell and Pimm (1976) also generally provides reasonable estimates for the interaction coefficients. However, each time the parametric regression coefficient is nonzero, the mean value of the SW coefficient is significantly different from expectation (P values, Table 2). Most of these differences are small, but in the case of consumer 2 and mutualism, the departure from expectation is large. In general, the conclusions drawn from individual data sets are correct. In the neutralism case, four to six type I errors occur at the .05 level of significance, and for mutualism and competition five or fewer type II errors occur except for MELS applied to the competition case for consumer 2.

Introducing an additional 15 unrelated variables increases the mean condition index for OLS; for MELS and SW the introduction of the nuisance variables has much less effect because the minor dimensions or extraneous variables are deleted from the analysis (Table

TABLE 3. Simulation summary with additional 15 nuisance variables. OLS = ordinary least squares; MELS = maximum eigenvalue least squares; SW = stepwise regression.

Statistics*	Response variable: Consumer 1 OLS	Consumer 1 MELS	Consumer 1 SW	Consumer 2 OLS	Consumer 2 MELS	Consumer 2 SW
Competition						
Parametric	−0.758	−0.758	−0.758	−1.259	−1.259	−1.259
Estimate	−0.772	−0.618	−0.669	−1.255	−0.100	−1.062
SE	0.009	0.034	0.009	0.014	0.061	0.024
C.Index	1278.4	33.9	28.6	1246.2	24.6	132.5
|Z|	1.57	4.15	9.89	0.27	19.12	8.14
P	.12	<.001	<.001	.79	<.001	<.001
Neutralism						
Parametric	0.000	0.000	0.000	0.000	0.000	0.000
Estimate	0.017	0.014	0.010	0.084	0.028	0.013
SE	0.008	0.008	0.006	0.041	0.022	0.013
C.Index	1199.4	21.4	35.6	1247.0	24.2	2.2
|Z|	2.06	1.81	1.54	2.07	1.32	1.02
P	.04	.07	.12	.04	.19	.31
Mutualism						
Parametric	0.438	0.438	0.438	2.276	2.276	2.276
Estimate	0.440	0.436	0.423	2.300	0.634	0.498
SE	0.005	0.005	0.006	0.028	0.062	0.021
C.Index	1413.6	21.3	38.5	1450.9	22.7	2.97
|Z|	0.32	0.32	2.69	0.85	26.4	85.1
P	.75	.75	.01	.40	<.001	<.001

*See Table 2.

3). The extra variables have little effect on the mean OLS coefficients or the inferential errors. However, the nuisance variables do increase the difference between the actual and estimated mean coefficients for consumer 2 when SW and especially MELS are used. The frequency of type II errors for MELS increases to unacceptable levels for competition, particularly for consumer 2.

DISCUSSION

The first simulation represented an interaction in which both consumers were correlated almost identically with the resources and, therefore, were positively correlated with each other (Table 1). The positive correlation between consumers was misleading, whereas the partial regression coefficients indicated a negative (competitive) relationship when adjustments were made for the resources (Table 2). The greatly reduced condition indices (Table 2) for the SW and MELS procedures indicate that the deleted variables and/or eigenvectors did reduce the collinearity among the resource predictors but had little impact on the collinearity involvement of the consumer predictor variable with the remaining resource predictor variables (Table 4). All three regression techniques gave consistent estimates of the interaction coefficient. The mean coefficients for MELS and SW were statistically different from expectation but not by enough to seriously bias conclusions drawn from the analyses. In addition, MELS also seemed to be slightly less powerful than the other techniques.

The second simulation shown in Table 2 represented a situation in which consumer 1 was correlated with the resources but consumer 2 was not (Table 1), and there was no correlation between consumers (neutralism). Again, SW and MELS procedures, by eliminating variables and/or eigenvectors, greatly reduced the condition indices (Table 2), suggesting a reduction in collinearity relative to OLS. When consumer 2 was the response variable, the SW procedure also dramatically reduced the collinearity involvement of the consumer predictor variable (Table 4) without distorting the interaction coefficient estimate (Table 2). All three techniques performed well, yielding between four and six type I errors at the .05 level of significance.

The third simulation presented in Table 2 was derived from a correlation matrix in which consumer 1 was correlated with the resources and consumer 2 was not, but consumers 1 and 2 showed a simple positive correlation with each other (Table 1). The partial regression coefficients (Table 2) showed this to be a mutualistic situation and, except for SW on consumer 2, the three regression techniques represented the parametric conditions reasonably well. Once again, the condition indices (Table 2) suggest that SW and MELS, through the elimination of variables and/or eigenvectors, reduced collinearity. As in the neutralism case for consumer 2, SW reduced the collinearity involvement of the consumer 1 predictor variable (Table 4) but now

TABLE 4. A summary of the coefficients of determination (R^2)* for the consumer predictor regressed on the remaining resource predictors. OLS = ordinary least squares; MELS = maximum eigenvalue least squares; SW = stepwise regression.

Predictor variable		OLS	MELS	SW
Without nuisance variables				
Consumer 2	Competition	0.72	0.70	0.67
	Neutralism	0.09	0.07	0.05
	Mutualism	0.11	0.07	0.07
Consumer 1	Competition	0.83	0.81	0.81
	Neutralism	0.84	0.80	0.16
	Mutualism	0.84	0.82	0.22
With nuisance variables				
Consumer 2	Competition	0.81	0.75	0.71
	Neutralism	0.39	0.37	0.08
	Mutualism	0.42	0.38	0.21
Consumer 1	Competition	0.88	0.49	0.82
	Neutralism	0.89	0.48	0.13
	Mutualism	0.89	0.37	0.24

* $R^2 = (1 - 1/q)$, where q = variance inflation factor (Gunst and Mason 1980) for the consumer predictor variable.

the excluded resource variables resulted in a distorted estimate of the interaction coefficient (Table 2). Apparently, the excluded variables had little direct effect on consumer 2 (the response variable), but without adjusting for their variation, the true relationship between consumers 1 and 2 was poorly estimated.

The regression methods generally resulted in reasonable mean values and standard errors for the regression interaction coefficients, which were supposed to estimate coefficients of interspecific interaction. This agreement with expectation is certainly true if one is only interested in the sign and not the magnitude of the coefficient (Pimm 1985). A large statistical literature has clearly demonstrated the negative consequences of collinearity, and the magnitude of the condition indices for OLS in Tables 2 and 3 indicates collinearity in the simulated data. However, the eigenstructure for the consumer predictor that was the focus of our attention (Table 5), indicates that the consumer predictor was not strongly associated with the dimension of the data most responsible for the collinearity problem (i.e., the consumer predictor did not load heavily on the eigenvector associated with the smallest eigenvalue). It can be shown (Ginevan and Carnes 1981) that these small eigenvalues are responsible for the variance inflation of regression coefficients. The consumer predictor variables' lack of association with the minor dimension of the predictor variable data structure is the reason we did not observe the collinearity effects we had initially anticipated. OLS would be expected to perform well in estimating the coefficients associated with noncollinear predictor variables. The condition index, which is a measure of the degree of collinearity in the entire data structure, is misleading when applied to a single coefficient in the regression equation. It does, however, provide a warning that collinearity problems may exist and caution should be applied when there is an interest in interpreting individual coefficients in the model.

Even though we were motivated by the controversy of the Crowell-Pimm approach to the estimation of competition coefficients, this paper really addresses two basic statistical issues. Collinearity, which we have discussed at length, is one issue and the effect of omitting predictor variables (misspecification bias) is the second issue. As our collinearity techniques clearly demonstrate, these two issues are interrelated. Even though the consumer predictor variable is not associated with the minor dimension of the data structure in any of our simulations (Table 5), the R^2s of Table 4 demonstrate that this variable was correlated with the other predictor variables (ranging from 0.05 to 0.89). The significant correlations observed for consumer 1 and the consumer 2 competition case are also suggested by the loadings of these variables on the eigenvector associated with the second smallest eigenvalue (compare Tables 4 and 5). Table 4 also demonstrates that SW and MELS performed as expected. Without exception, the correlation (R^2) of the consumer predictor with the resource predictors was reduced when these techniques were applied. More important, Table 4 focuses our attention on those cases where there is a potential cost in the use of the collinearity techniques. When the R^2 drops for SW or MELS relative to OLS because of the elimination of variables and/or eigenvectors (Table 4), what is the effect on the estimate of the interaction coefficient? Since all the variables used to generate the simulated data are included in it, the OLS model is always correctly specified and is used as a reference for comparing the effects of misspecification on the SW and MELS solutions.

In our simulations where collinearity existed but the consumer predictor was not primarily involved in the collinearity, the OLS procedure outperformed the techniques designed for dealing with collinearity. This outcome emphasizes the potential cost (misspecification bias) associated with a reduction of collinearity. The collinearity techniques reduced correlations among the set of predictor variables as indicated by the smaller condition indices; however, information is always lost when either variables (SW) or eigenvectors (MELS) are deleted from the model. This information loss may bias the resulting estimates of interspecific interactions. For example, in the SW mutualism case with consumer 2 as the response (Table 2 or 3), consumer 1 (the predictor) is collinear with the resource variables and associated with a small eigenvalue (Table 5). In this case, we see that the R^2 for the consumer predictor regressed on the resource predictors remaining in the model dropped appreciably (Table 4). The condition of the design matrix (an index of collinearity) was greatly improved by the deletion of predictor variable(s), but

TABLE 5. Eigenvalues (λ) and associated eigenvector loadings for the consumer predictor variable in the simulations without nuisance variables.

	Eigenvalue and loading	Eigenvectors				
		1	2	3	4	5
Invariant*	λ	2.852	1.625	0.430	0.088	0.004
	Consumer 1	0.300	0.646	−0.256	0.654	−0.014
Competition	λ	2.834	1.570	0.437	0.155	0.004
	Consumer 2	0.280	0.656	−0.305	0.631	0.024
Neutralism	λ	2.698	1.000	0.926	0.372	0.005
	Consumer 2	0.000	1.000	0.000	0.000	0.000
Mutualism	λ	2.698	1.042	0.885	0.371	0.004
	Consumer 2	0.014	0.859	0.509	−0.049	0.009

* Invariant eigenstructure when consumer 2 is the response variable.

these variable(s) were necessary for the accurate estimation of the interaction coefficient (Table 2 or 3).

It is important to emphasize that the collinearity techniques described here are concerned with the collinearity structure of the predictor variables and do not directly involve the response variable. In SW, a variable highly correlated with the response variable is not able to enter the model if it also happens to be highly correlated with a predictor variable already in the model. MELS deals only with the eigenstructure of the predictor space. A selection criterion for eigenvectors that are correlated with the response variable does exist in principal components regression. However, if the response variable is associated with a minor dimension (i.e., an eigenvector associated with a small eigenvalue) of the data, this approach will not reduce the variance inflation problem associated with collinearity, which was the motivation for using the technique. In the present study, when the response variable was not strongly correlated with a minor dimension or the deleted variable was unessential, the MELS and SW approaches not only reduced collinearity, but also provided reasonable estimates of the interaction coefficients. Given these results, we would be forced to conclude that (at least in the simple situation presented here) the biased regression techniques worked well, but one should be careful when eliminating either eigenvectors or predictor variables. This care should be exercised even when the predictor variables have no significant influence on the response variable.

This tentative conclusion is based on the inclusion of at most four variables that have little effect on the dependent variable. Our second set of simulations (summarized in Table 3) was designed to test if this robustness extended to more nuisance variables. Inclusion of 15 extraneous variables did not seem to change estimation when OLS or SW was used, but the MELS coefficients were more biased. The inclusion of the extraneous variables increased the probability of predictive information being associated with small eigenvectors, which were discarded by the MELS approach.

Given the success of regression with these particular data, how can we explain the results of Abramsky et al. (1986), cited as in preparation by Pimm (1985), which clearly demonstrate the failure of regression coefficients to detect competition when it could be shown to exist experimentally? One possibility, suggested by Pimm (1985) and discussed by Abramsky et al. (1986), is that the interactions involved are nonlinear, which makes extrapolation from near-equilibrium conditions to complete removal of a species unreliable. This study suggests a second possibility. Our simulations represented a situation in which all influential variables were measured (i.e., values for consumer densities were generated by using only those variables included as candidates for entry into a regression model). Some of our predictor variables were not influential under certain correlation structures, but all relevant variables were measured in all cases. With actual field data, there will always be more environmental variables than can be measured, and one can only hope to measure those with greatest influence. Similarly, there are more biotic variables involved than can be measured, so attention is focused on those species thought to be most influential or of particular interest. With field data, the possibility that the influence of an unmeasured variable on both potential competitors will be interpreted as an interspecific interaction always exists. In fact, we see a strong connection between the suggestion of Abramsky et al. (1985) that strong habitat affinities, which were not measured, could result in a positive correlation between densities of competing consumers (our first simulation) and the phenomenon of apparent competition (Holt 1977). Apparent competition arises when an uncensused predator or parasite causes negatively correlated population fluctuations between consumers when no competition actually exists. In each instance, a critical element of the resource–consumer matrix is not measured.

The potential bias in coefficient estimation caused by the omission of a variable can be demonstrated by manipulating the normal equations for multiple regression to show the relationship that exists between simple and multiple regression (Wonnacott and Wonnacott 1979). In the simple two predictor variable case

$$B_{Y1} = B_1 + B_2 \cdot B_{21} \quad (2)$$

the simple regression coefficient of Y on predictor 1 (B_{Y1}) is seen to be a function of the multiple regression coefficient of Y on predictor 1 (B_1) and an indirect product term involving the multiple regression coefficient of Y on predictor 2 (B_2) and the simple regression coefficient of predictor 2 on predictor 1 (B_{21}). It is possible for the indirect relation to determine the sign and magnitude of the simple regression coefficient (B_{Y1}), and bias will occur if these effects are not measured or are omitted.

Even when all elements of a community have been measured, the possibility of misinterpretation still exists with use of ordinary stepwise regression in which measures of resources and other consumers are simultaneously considered as candidates for entry in the model. Suppose that, as in our first simulation, two consumer species respond positively to the same three or four elements of habitat structure and negatively to each other. The initial correlation between consumers may well be positive since each reflects the cumulative effects of the habitat variables. This correlation could easily be the most important single influence on abundance of the consumers because any single habitat variable represents only a fraction of the habitat information. In an ordinary stepwise procedure, the early entry of the consumer variable might preclude entry of any single habitat variable. Stepwise procedures that use forward selection are particularly sensitive to this problem.

It should be noted at this point that there are alternative regression strategies to those described here (e.g., latent root regression, stagewise fitting, SW regression using backward elimination, and best subset selection regression) that may not be as strongly influenced by the correlation structure among predictor variables. In this study, OLS was selected as a reference solution; SW was chosen because we were interested in the discrepancies reported for the Crowell-Pimm approach, and MELS was chosen because we thought it would outperform the other approaches.

The modified approach of Crowell and Pimm (1976) partially circumvents the variable selection difficulties by first conducting the stepwise procedure only on the habitat variables and then forcing the consumer variables into the model. However, as our simulation demonstrates, even this approach can lead to distorted interaction estimates. If one forces in all variables or uses OLS or MELS, the true picture may be revealed, but one also risks computational problems with OLS when using correlation matrices that tend toward singularity (i.e., large condition index) and distortion problems with MELS when the response variable (consumer density) is highly correlated with unstable minor dimensions of the data.

Identification and demonstration of the potential problems in the application of regression techniques are straightforward. Unfortunately, simple solutions to these problems do not exist. We have discussed the potential bias that can occur by the deletion or omission of important predictor variables. This might suggest that one should measure and use as many predictor variables as possible. However, overspecified models have the same bias problem as underspecified models and increase the probability of introducing redundancies (collinearity) into the data structure. Uninformative predictor variables also reduce the degrees of freedom available for estimating the error term without significantly reducing the error term itself, which leads to less sensitive statistical tests. Therefore, variables should not be considered just because they are available but instead should be selected as regression candidates for theoretical reasons. It should also be noted that a completely misspecified model can have a perfectly acceptable coefficient of multiple determination.

After a set of predictor variables has been selected, attention should be focused on the structure of these variables. The assumption that the predictor variables are nonstochastic and measured without error is always violated in ecological studies. Thus, the precision of measurement for a potential predictor might be one criterion used in an initial screening of predictor candidates. The extraction of eigenvalues and eigenvectors (principal components [PC] analysis) provides essential information concerning interrelationships among the predictor variables and is the multivariate analog of graphing data. PC analysis should be an early step in data exploration and is one of the reasons we were interested in the performance of MELS, which emphasizes the eigenstructure of the data.

Collinearity is emphasized in this paper not only because of its inherent role in ecological data but also because it can occur for so many different reasons. For example, the year × species interaction terms in the Crowell and Pimm approach (1976) are products of variables already in the model. These cross-product terms will introduce collinearity and potential bias problems unless the design is balanced (i.e., same relative frequency of each species in every year). Collinearity can also be induced by an outlier (extreme observation) in the data set. If the collinearity is sampling induced, it may be necessary to collect additional data. Choosing these new data in the direction of the minor dimension(s) (i.e., eigenvectors) will improve the precision of estimation and reduce collinearity (Silvey 1969). Finally, collinearity may occur simply because the predictors actually do share information. Perhaps correlations that are likely to arise from field studies may not be large enough to cause severe collinearity problems but without an initial PC analysis, this cannot be known. It is also likely that field data will have lower coefficients of determination between response and predictor variables that will increase statistical uncertainty.

The next step in the analysis of the data is model

building. It should now be clear that each regression technique we have discussed should not be applied mechanically. In the absence of a theoretical framework, model building is an exploratory exercise. This suggests that many approaches should be applied (e.g., best subset and forward as well as backward selection techniques). Being aware of the limitations of these approaches also suggests precautions that should be taken in their application. Influence diagnostics (Belsley et al. 1980), which are widely available in software packages (e.g., BMDP, SAS), should routinely be used to identify data points that influence the regression coefficient of interest. Partial residual plots (Gunst and Mason 1980) provide simple graphical indications of model misspecification and the proper functional form (i.e., transformation) for predictor variables.

It should be emphasized that we were not attempting to investigate pathological examples of collinearity, nor were we exploring a broad range of correlation structures. Instead, our correlation structures were nonrandomly selected to represent data that could actually occur in field studies. The correlation structure among resource dimensions was invariant, as was the correlation structure of consumer 1 with the resource dimensions. Within these constraints, we manipulated the correlation structure of the second consumer to achieve the three types of consumer interactions presented in this paper.

In summary, the problems of misspecification bias and collinearity are interrelated and their solutions are antagonistic. Many approaches can be used to address the problem of collinearity but none of the approaches discussed in this paper can protect against the potential bias created by unmeasured or deleted predictor variables. Nevertheless, the results from this simulation study with its restricted correlation structure and specific collinearity associations support the contention of Pimm (1985) that competition coefficients can be estimated at least as to sign if not magnitude. Therefore, when carefully applied, the approach of Crowell and Pimm (1976) can provide a valuable tool in the investigation of competition. Finally, we echo the comment of Abramsky et al. (1986) that "we hope that the results of this study will encourage other ecologists to subject the method to additional field tests before using it in the future."

Acknowledgments

I would like to give special thanks to Mike Miller at Argonne National Laboratory for his thoughtful reviews and advice. In addition, we would like to acknowledge the constructive criticisms of the reviewers, which have led to a much improved manuscript. Work supported by the United States Department of Energy, Office of Health and Environmental Research. The submitted manuscript has been authored by a contractor of the U.S. Government under contract No. W-31-109-ENG-38. Accordingly, the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes.

Literature Cited

Abramsky, Z., M. A. Bowers, and M. L. Rosenzweig. 1986. Detecting interspecific competition in the field: testing the regression method. Oikos **47**:199–204.

Abramsky, Z., M. L. Rosenzweig, and S. Brand. 1985. Habitat selection in Israel desert rodents: comparison of a traditional and a new method of analysis. Oikos **45**:79–88.

Afifi, A. A., and V. Clark. 1984. Computer-aided multivariate analysis. Lifetime Learning Publications, Belmont, California, USA.

Belsley, D. A., E. Kuh, and R. E. Welsch. 1980. Regression diagnostics: identifying influential data and sources of collinearity. John Wiley and Sons, New York, New York, USA.

Carnes, B. A. 1980. Habitat selection in a prairie rodent community. Dissertation. University of Kansas, Lawrence, Kansas, USA.

Crowell, K. L. 1983. Islands—insight or artifact? Population dynamics and habitat utilization in insular rodents. Oikos **41**:442–454.

Crowell, K. L., and S. L. Pimm. 1976. Competition and niche shifts of mice introduced onto small islands. Oikos **27**:251–258.

Dueser, R. D., and J. G. Hallett. 1980. Competition and habitat selection in a forest-floor small mammal fauna. Oikos **35**:293–297.

Farrar, D. E., and R. R. Glauber. 1967. Multicollinearity in regression analysis: the problem revisited. Review of Economics and Statistics **49**:92–107.

Ginevan, M. E., and B. A. Carnes. 1981. Approaches to problems of collinearity and dimensionality in studies of disease-environment association. Proceedings of the DOE Statistical Symposium, Brookhaven National Laboratory, Long Island, New York, USA.

Green, P. E., and J. D. Carroll. 1976. Mathematical tools for applied multivariate analysis. Academic Press, New York, New York, USA.

Greenberg, E. 1975. Minimum variance properties of principal component regression. Journal of the American Statistical Association **70**:194–197.

Gunst, R. F., and R. L. Mason. 1977. Biased estimation in regression: an evaluation using mean squared error. Journal of the American Statistical Association **72**:616–628.

Gunst, R. F., and R. L. Mason. 1980. Regression analysis and its application: a data-oriented approach. Marcel Dekker, New York, New York, USA.

Hallett, J. G. 1982. Habitat selection and the community matrix of desert small-mammal fauna. Ecology **63**:1400–1410.

Hallett, J. G., M. A. O'Connell, and R. L. Honeycutt. 1983. Competition and habitat selection: test of a theory using small mammals. Oikos **40**:175–181.

Hallett, J. G., and S. L. Pimm. 1979. Direct estimation of competition. American Naturalist **113**:593–600.

Hawkins, D. M. 1973. On the investigation of alternative regressions by principal component analysis. Applied Statistics **22**:275–286.

Holt, R. D. 1977. Predation, apparent competition and the structure of prey communities. Theoretical Population Biology **12**:197–229.

Levins, R. 1968. Evolution in changing environments. Princeton University Press, Princeton, New Jersey, USA.

MacArthur, R. H., and R. Levins. 1967. The limiting similarity, convergence, and divergence of coexisting species. American Naturalist **101**:377–385.

Marquardt, D. W. 1970. Generalized inverses, ridge regres-

sion, biased linear estimation, and nonlinear estimation. Technometrics **12**:591–612.

Mason, R. L., R. F. Gunst, and J. T. Webster. 1975. Regression analysis and problems of multicollinearity. Communications in Statistics **4**:277–292.

Massy, W. F. 1965. Principal components regression in exploratory statistical research. Journal of the American Statistical Association **60**:234–256.

McDonald, G. C., and R. Schwing. 1973. Instabilities of regression estimates relating air pollution to mortality. Technometrics **15**:463–481.

Odum, E. P. 1971. Fundmentals of ecology. W. B. Saunders, Philadelphia, Pennsylvania, USA.

Pimm, S. L. 1985. Estimating competition coefficients from census data. Oecologia (Berlin) **67**:588–590.

Porter, J. H., and R. D. Dueser. 1982. Niche overlap and competition in an insular small mammal fauna: a test of the niche overlap hypothesis. Oikos **39**:228–236.

Rosenzweig, M. L., Z. Abramsky, and S. Brand. 1984. Estimating species interactions in heterogeneous environments. Oikos **43**:329–340.

Rosenzweig, M. L., Z. Abramsky, B. Kotler, and W. Mitchell. 1985. Can interaction coefficients be determined from census data? Oecologia (Berlin) **66**:194–198.

SAS. 1985. SAS user's guide: statistics. 1982 edition. SAS Institute, Cary, North Carolina, USA.

Schoener, T. W. 1974. Competition and the form of habitat shift. Theoretical Population Biology **6**:265–307.

———. 1985. On the degree of consistency expected when different methods are used to estimate competition coefficients from census data. Oecologia (Berlin) **67**:591–592.

Seifert, R. P., and F. H. Seifert. 1976. A community matrix analysis of *Heliconia* insect communities. American Naturalist **110**:461–483.

Silvey, S. D. 1969. Multicollinearity and imprecise estimation. Journal of the Royal Statistical Society **B 31**:539–552.

Sokal, R. R., and F. J. Rohlf. 1969. Biometry. W. H. Freeman, San Francisco, California, USA.

Wilson, B. K. 1985. Simultaneity and its impact on ecological regression applications. Biometrics **41**:435–445.

Wonnacott, R. J., and T. H. Wonnacott. 1979. Econometrics. John Wiley and Sons, New York, New York, USA.